冈比亚按蚊犬尿氨酸甲酰胺酶抑制剂的虚拟筛选

宋帅葆[1]，艾上杰[2]，关 怀[1,*]，韩 谦[1,*]

(1. 海南大学热带农林学院，海口 570100；2. 海南大学信息科学技术学院，海口 570100)

摘要：【目的】利用计算机模拟技术，对冈比亚按蚊 *Anopheles gambiae* 犬尿氨酸甲酰胺酶(kynurenine formamidase, KFase)的潜在抑制剂进行虚拟筛选，以获得可以削弱冈比亚按蚊作为中间宿主传播疟疾等蚊媒疾病的候选杀蚊剂。【方法】下载冈比亚按蚊 KFase 的氨基酸序列，通过 BLAST 方法查询不同物种中的同源蛋白质，并利用 MEGA6 最大似然法(maximum likelihood method)构建进化树，选择适于作为模板的同源蛋白黑腹果蝇 *Drosophila melanogaster* KFase 晶体结构(PDB ID: 4E14)，对冈比亚按蚊 KFase 进行三维建模。利用随机森林算法对小分子化合物数据库进行筛选，并对筛选结果进行处理，模拟自然条件下有机小分子与冈比亚按蚊 KFase 的结合以及分子对接，从而筛选出冈比亚按蚊 KFase 的潜在抑制剂。【结果】获得 3 个小分子化合物与冈比亚按蚊 KFase 结合的亲和能较低，分别是：N-(2,4-diketo-1H-pyrimidin-6-yl)-2-fluoro-benzamide；3-(4-fluorophenyl)-2, 4-dioxo-1, 2, 3, 4-tetrahydropyrimidine-5-carboxylic acid；N-(2-oxo-2, 3-dihydro-1H-imidazo[4,5-b] pyridin-5-yl)-succinamic acid。它们与冈比亚按蚊 KFase 结合的亲和能分别为：–9.0, –8.7 和 –8.9 kcal/mol。【结论】N-(2,4-diketo-1H-pyrimidin-6-yl)-2-fluoro-benzamide, N-(2-oxo-2,3-dihydro-1H-imidazo[4,5-b] pyridin-5-yl)-succinamic acid 和 3-(4-fluorophenyl)-2,4-dioxo-1, 2,3,4-tetrahydropyrimidine-5-carboxylic acid 是冈比亚按蚊犬尿氨酸甲酰胺酶的潜在竞争性抑制剂，这些化合物是否可作为杀蚊剂的候选化合物有待实验验证。

关键词： 冈比亚按蚊；疟疾；犬尿氨酸甲酰胺酶；抑制剂；虚拟筛选；随机森林；生物信息学

中图分类号： Q966 **文献标识码：** A **文章编号：** 0454-6296(2018)01-0068-11

Virtual screening of inhibitors for kynurenine formamidase of *Anopheles gambiae* (Diptera: Culicidae)

SONG Shuai-Bao[1], AI Shang-Jie[2], GUAN Huai[1,*], HAN Qian[1,*] (1. Institute of Tropical Agriculture and Forestry, Hainan University, Haikou 570100, China; 2. College of Information Science & Technology, Hainan University, Haikou 570100, China)

Abstract: 【Aim】To screen potential inhibitors targeting kynurenine formamidase (KFase) of *Anopheles gambiae*, which could be a candidate insecticide to reduce malaria transmission, by virtual screening. 【Methods】Protein sequences homologous with KFase of *An. gambiae* were searched and downloaded from NCBI using BLASTP web server. Phylogenetic tree of homologous proteins was constructed by MEGA6 using maximum likelihood method. Homology modeling of KFase of *An. gambiae* was performed by SWISS-MODEL web server using KFase of *Drosophila melanogaster* (PDB ID: 4E14) as a template. The small-molecule compounds were downloaded from ZINC database and then screened by the method of random forest. Docking analysis of the homology model and selected small molecule compounds was carried out, and the screening results were further validated using molecular dynamics simulation.

基金项目：国家自然科学基金项目(31472186)；海南大学科研专项(hdkytg201702)

作者简介：宋帅葆，男，1990 年 4 月生，河南南阳人，硕士研究生，研究方向为野生动植物保护与利用，E-mail: 810157048@qq.com

* 通讯作者 Corresponding authors, E-mail: qianhan@hainu.edu.cn; huai_guan@163.com

收稿日期 Received: 2017-06-05；接受日期 Accepted: 2017-08-14

【Results】 Three small-molecule compounds, *i. e.*, N-(2, 4-diketo-1H-pyrimidin-6-yl)-2-fluorobenzamide, 3-(4-fluorophenyl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylic acid, and N-(2-oxo-2, 3-dihydro-1H-imidazo[4, 5-b] pyridin-5-yl)-succinamic acid, have the lowest docking energy to KFase of *An. gambiae* with the affinity energy of −9.0, −8.7 and −8.9 kcal/mol, respectively. 【Conclusion】The three small-molecule compounds, *i. e.*, N-(2,4-diketo-1H-pyrimidin-6-yl)-2-fluorobenzamide, 3-(4-fluorophenyl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylic acid, and N-(2-oxo-2, 3-dihydro-1H-imidazo[4, 5-b] pyridin-5-yl)-succinamic acid, could be candidate competitive inhibitors of KFase of *An. gambiae*. Whether these compounds can be used as candidate mosquitocides needs further experimental validation.

Key words: *Anopheles gambiae*; malaria; kynurenine formamidase; inhibitor; virtual screening; random forest; bioinformatics

疟疾(malaria)是由疟原虫 *Plasmodium* spp. 导致的一种传染性疾病,其传播媒介包括冈比亚按蚊 *Anopheles gambiae* 等多种按蚊属 *Anopheles* 蚊虫。据估计,2015 年全球共有 2.14 亿疟疾新增病例,大约 43.8 万人死于疟疾,而其中有 80% 感染人数和 78% 的死亡人数在非洲(WHO, 2015)。冈比亚按蚊 *An. gambiae* 属于动物界昆虫纲双翅目蚊科按蚊属,主要分布于撒哈拉以南非洲地区,由于其传播包括疟疾在内的多种疾病,因此又称非洲疟蚊。冈比亚按蚊的代谢影响在其体内的疟原虫的生长发育,如色氨酸的代谢产物黄尿酸是疟原虫小配子发生所需的信号分子(Garcia *et al.*, 1998)。

在大多数真核生物体内,色氨酸代谢主要有两条途径,其中一条沿 5-羟色胺的代谢途径,另外一条沿犬尿氨酸的代谢途径。犬尿氨酸代谢途径也存在于包括冈比亚按蚊在内的按蚊体内。犬尿氨酸途径的一些中间产物有着特殊的生物活性,与一些疾病的发生有密切关系,除疟疾外,还有癌症、精神分裂和神经退行性疾病等(Han *et al.*, 2012)。蚊的犬尿氨酸途径中,色氨酸逐步代谢成犬尿氨酸,犬尿氨酸可转化为黄尿酸。犬尿氨酸甲酰胺酶(kynurenine formamidase, KFase),也称芳基甲酰胺酶(arylformamidase),该酶能够可逆地催化 N-甲酰-L-犬尿氨酸(N-formyl-L-kynurenine, NFK)水解,产生 L-犬尿氨酸及甲酸。该酶是色氨酸代谢的犬尿氨酸途径中第 2 个酶,是犬尿氨酸途径的限速酶之一。抑制 KFase 不仅会对蚊虫的正常生理功能造成不利影响,还可以减少黄尿酸的生成,抑制疟原虫小配子的生成,使按蚊体内的疟原虫无法受精产生卵囊及子孢子,从而抑制疟原虫的传播。

冈比亚按蚊作为寄生虫学研究的模式生物之一,具备寄生虫传播媒介的典型特征。因此研究冈比亚按蚊 KFase 的抑制剂,有利于蚊虫防控技术的发展。本研究的目的是,利用在线数据库、分子动力学模拟(molecular dynamics simulation, MD)及分子对接等方法,对冈比亚按蚊 KFase 进行三维结构建模,并依据模型对其潜在抑制剂进行虚拟筛选,筛选结果可用于新型杀蚊剂的进一步实验室研究,降低易感人群感染疟疾等虫媒传染病的风险。

1 材料与方法

1.1 目标氨基酸序列

UniProt 数据库(www.uniprot.org)收录了冈比亚按蚊 KFase 的氨基酸序列,其 UniProt 编号为 Q7QKH2,与其序列同时得到确认的还有其活性位点。下载该序列,作为同源结构建模的目标氨基酸序列。

1.2 三维结构模板选择与同源建模

使用 NCBI(National Center for Biotechnology Information)提供的 BLASTP(blast.ncbi.nlm.nih.gov)在无冗余蛋白序列数据库(non-redundant protein sequence database)中搜索同源序列并下载比对序列的 FASTA 文件,利用 MEGA6 构建进化树(Tamura *et al.*, 2013),先使用 ClustalW 比对序列,最大似然法(maximum likelihood, ML)构建系统发育树,Bootstrap 设置为 500。选择进化距离最近并且在 PDB 数据库(www.rcsb.org)中有三维结构的同源蛋白作为同源建模所使用的结构模板,在 PDB 数据库(www.rcsb.org)中下载该结构模板,利用 AutoDock Tools1.5.6 (Morris *et al.*, 2009)对其进行编辑,删除其中的水分子及其他小分子,作为同源建模结构模板。

将目标氨基酸序列(UniProt ID: Q7QKH2)与结

构模板输入 SWISS-MODEL 服务器（www.swissmodel.expasy.org）进行同源建模（Arnold *et al*., 2006；Kiefer *et al*., 2009；Biasini *et al*., 2014），获得的结构模型用于作为虚拟筛选中的受体模型。

1.3 动力学模拟与建模结果的评估

冈比亚按蚊 KFase 模型的动力学模拟是通过 Gromacs 5.1.2（Abraham *et al*., 2015；Kutzner *et al*., 2015）运算完成的。动力学模拟所使用的力场是 Gromacs5.1.2 内置的 Gromos96 力场，水体设置为立方型（Cube），设置周期边界，在水体中添加钠离子和氯离子确保体系呈电中性。随后进行能量最小化、恒定粒子数、体积和温度（constant number of particles, volume, and temperature, NVT）平衡，以及恒定粒子数、压力和温度（constant number of particles, pressure, and temperature, NPT）平衡，然后模拟冈比亚按蚊 KFase 在水环境中 50 ns 的动力学过程。最后利用 Gromacs5.1.2 自带的分析工具获取数据，并利用 Xmgrace（http：// plasma-gate.weizmann.ac.il/Grace）对产生的数据进行绘图，根据动力学模拟过程中体系能量对建模得到的 KFase 结构模型进行稳定性评估。

1.4 蛋白质与底物复合物的动力学模拟

利用 AutoDock Vina（Trott and Olson, 2010）和 AutoDock Tools1.5.6 对 KFase 酶蛋白与本酶目标底物 N-甲酰-L 犬尿氨酸（N-formal-L-kynurenine, NFK），在酶蛋白的活性中心附近进行刚性对接，酶蛋白作为受体设置成完全刚性，而 NFK 模型则将尽可能多的共价键设置成可旋转，保证本分子模型的柔性尽可能地高。通过本次刚性分子对接，得到对接结果的 PDBQT 文件，从对接结果的 PDBQT 文件中选择出两个分子结合得最佳的构象，即配体中的酰胺键与酶蛋白活性中心的丝氨酸残基距离尽可能接近，同时亲和能还要尽可能地低。

利用 AutoDock Tools1.5.6 分别提取受体冈比亚按蚊 KFase 和配体 NFK 的 PDB 格式分子模型，其中配体分子模型在保存前添加氢原子。利用 Ambertools16（Wang *et al*., 2006）软件及 ACPYPE（Silva and Vranken, 2012）软件生成能够适应 Amber 力场的、较为准确的小分子拓扑。按照 Gromacs5.1.2 复合物动力学模拟的方式定义水盒参数并填充水分子、添加钠离子和氯离子，并进行能量最小化、NVT 平衡和 NPT 平衡以及动力学模拟，模拟结束后提取冈比亚按蚊 KFase 的模型并保存为 PDB 格式，用于小分子化合物的虚拟筛选。

1.5 随机森林方法对小分子化合物数据库的筛选与评估

将氨基甲酸酯类化合物中一些具有代表性的有效物质选取出来，其中包括涕灭威（aldicarb）、西维因（carbaryl）、呋喃丹（carbofuran）、异索威（isolan）、异丙威（isoprocarb）、速灭威（metolcarb）、抗蚜威（pirimicarb）、氨基甲酸甲酯（urethylane, 又名尿基烷）（刘毅锋等, 2002），并加上 N-甲酰-D-犬尿氨酸（Nagai *et al*., 1994），将这 9 种物质组成一个集合，作为有效的阳性对照集合。

另收集其他 10 种相对分子质量接近、分子结构与上述几种物质包含一定相似性但无明显抑制作用的小分子有机物作为阴性对照集合。

将两个集合组成一个训练集，采用 MOLD2（Hong *et al*., 2008）软件，结合随机森林方法（Breiman, 2001）进行虚拟筛选。

将训练集中的阳性和阴性集合分别输入软件中，并对分子描述符进行评估和筛选，从 779 个分子描述符中选出对于筛选最有价值的描述符：

分子模型的 SDF 文件输入 MOLD2 软件后，首先去除每一列包含 0 的个数大于 85% 的描述符，然后去除方差小于 0.05 的描述符；最后去除绝对相关系数大于 0.95 的两个描述符中的一个。经过本轮筛选，描述符数量减为 171 个，随后利用特征选择进行下一轮的描述符的评估和筛选（谢娟英等, 2010）。

Chen 和 Lin（2006）给定训练样本第 i 个特征的 F-score 定义为：

$$F_i = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n+}(x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n-}(x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2}。$$

其中，正类和负类的样本数分别为 n_+ 和 n_-；$\bar{x}_i$，$\bar{x}_i^{(+)}$ 和 $\bar{x}_i^{(-)}$ 分别为第 i 个特征在整个数据集上的平均值，在正类数据集上的平均值和在负类数据集上的平均值；$x_{k,i}^{(+)}$ 为第 k 个正类样本点的第 i 个特征的特征值；$x_{k,i}^{(-)}$ 为第 k 个负类样本点的第 i 个特征的特征值；F 值越大，此特征的辨别能力越强。

基于 F-score 方法设计一个对分子描述符集合的评估算法，流程如下：

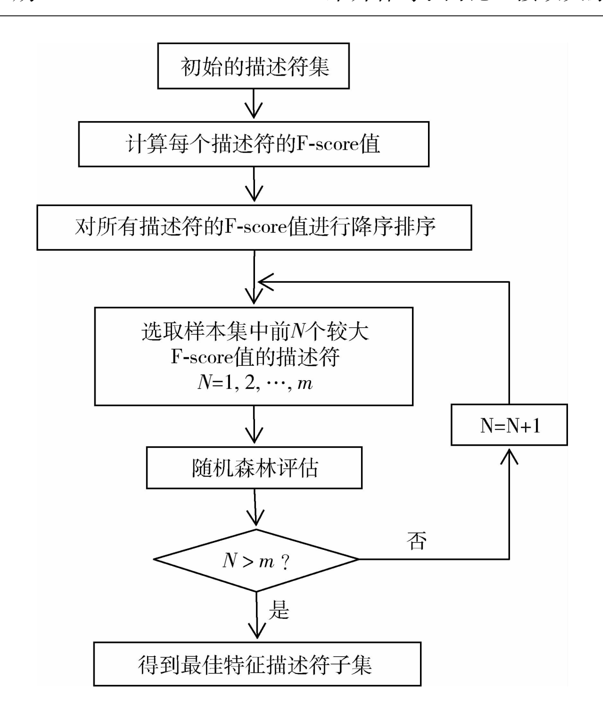


图 1　最佳分子描述符特征子集流程图

Fig. 1　Flow chart of characteristic subset of optimum molecular descriptors

经过此算法筛选，最终得到一定数目的最佳特征描述符子集。

把训练集的 19 个分子通过最佳特征描述符子集的 SOM 计算聚类。然后通过聚类来挑取训练集和测试集。

编订了 4 个神经元：

1 号神经元 6 个数据，序号为：[11 14 15 16 17 18]；2 号神经元 4 个数据，序号为：[10 12 13 19]；3 号神经元 4 个数据，序号为：[1 3 7 9]；4 号神经元 5 个数据，序号为：[2 4 5 6 8]。

随机森林是一个包含多个决策树的分类器，这些决策树是随机产生的，随机森林中决策树之间没有关联。

测试数据进入随机森林，每一棵决策树对测试数据进行分类，最后取所有决策树中分类结果最多的那类为最终的结果。默认的决策树 500 棵。

对于训练集的 19 个分子，在 4 组聚类中，分别选取 75% 为训练集，剩下的 25% 为测试集，检验随机森林筛选的有效性。

从 ZINC 数据库（http://zinc.docking.org/）下载 Frag Now 数据库文件，该文件中包含有超过 704 040个相对分子质量不大于 250 的化合物数据。将该文件拆分成多个 sdf 文件，并输入 MOLD2。利用前文选定的方法进行筛选，利用决策树进行投票。取大于 400 投票的数据，计算 5 次，取交集。最终得到筛选后的小分子化合物集合。

1.6　抑制剂筛选结果的分子对接

对冈比亚按蚊 KFase 与 NFK 复合物的动力学模拟结果进行整理，提取其中的冈比亚按蚊 KFase 结构。利用 AutoDock Vina（Tanchuk *et al*., 2015；Jaghoori *et al*., 2016）对此冈比亚按蚊 KFase 结构与 NFK 进行 100 次半柔性对接，了解复合物中冈比亚按蚊 KFase 与 NFK 的亲和能情况，作为对抑制剂筛选结果进行评估的参照数据。

利用 AutoDock Vina 对筛选后的小分子化合物进行半柔性对接。受体选用 KFase 与 NFK 复合物动力学模拟结果中的蛋白质结构，AutoDock Tools 中网格框体设置于活性中心附近并包含整个活性中心。

每个化合物对接 3 次，取亲和能最低的数据及对应构象，进行整理和排序，取排名靠前的化合物作为配体与酶形成复合物进行动力学模拟。

1.7　配体与酶复合物的动力学模拟

以 1.6 节获得的 3 种亲和能最低的化合物作为配体，结合前文 1.4 节中提及的蛋白质-底物复合物动力学模拟的方法，利用 Gromacs 5.1.2 逐个进行与冈比亚按蚊 KFase 的复合物动力学模拟，并结合 AutoDock Vina 的进行刚性对接最终确认其亲和能，与冈比亚按蚊 KFase 和 NFK 亲和能数据进行对比，确认其相对于 NFK 是否有足够强的竞争性。

2　结果

2.1　KFase 系统发育分析与模板的选择

基于 NCBI 提供的 BLASTP 对从 Uniprot 获得的冈比亚按蚊 KFase 的 FASTA 序列，在无冗余蛋白序列（non-redundant protein sequences）数据库中进行同源序列搜索。结果表明，冈比亚按蚊 KFase 与其他蚊虫的 KFase 最接近。在搜索到的同源序列中，仅黑腹果蝇 *Drosophila melanogaster* KFase 具有三维结构（www.rcsb.org，PDB ID：4E14）。冈比亚按蚊 KFase 与黑腹果蝇 KFase 的氨基酸序列一致性为 36%，活性中心的关键残基附近序列高度相似（图 2）。

利用 MEGA6 中 ClustalW 对同源蛋白氨基酸序列的全长进行比对，并用最大似然法（maximum likelihood，ML）建立进化树（图 3）。在分析中，Bootstrap 检验次数设为 500 次，并选择 LG + G 模型分析。从图 3 中也可以看到，在进化树中，冈比亚按蚊 KFase 和黑腹果蝇 KFase 进化距离不远。鉴于以

上原因，我们选择了黑复果蝇 KFase 的三维结构作为同源建模的结构模版。

2.2 冈比亚按蚊 KFase 的分子建模

经过 SWISS-MODEL 建模显示（图 4），目标氨基酸序列（UniProt ID：Q7QKH2）与结构模板（PDB ID：4E14）序列一致性 37.96%，使用 RAMPAGE（http：// mordred. bioc. cam. ac. uk/ ~ rapper/ rampage. php）分析 SWISS-MODEL 构建模型的拉氏构象图，结果允许区 93.3%，最大允许区 4.6%，不允许区 6%，模型结构合理。

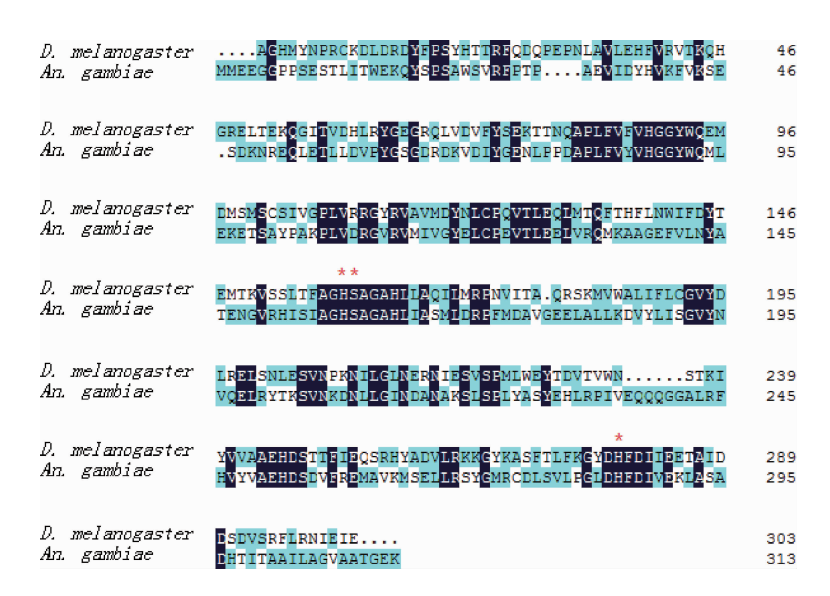


图 2 冈比亚按蚊与黑腹果蝇 KFase 的序列比对

Fig. 2 Amino acid sequence alignment of KFase proteins of *Anopheles gambiae* and *Drosophila melanogaster*

红色星号标注为催化活性中心周围残基。The residues around the active center are indicated by an asterisk and shown in red color.

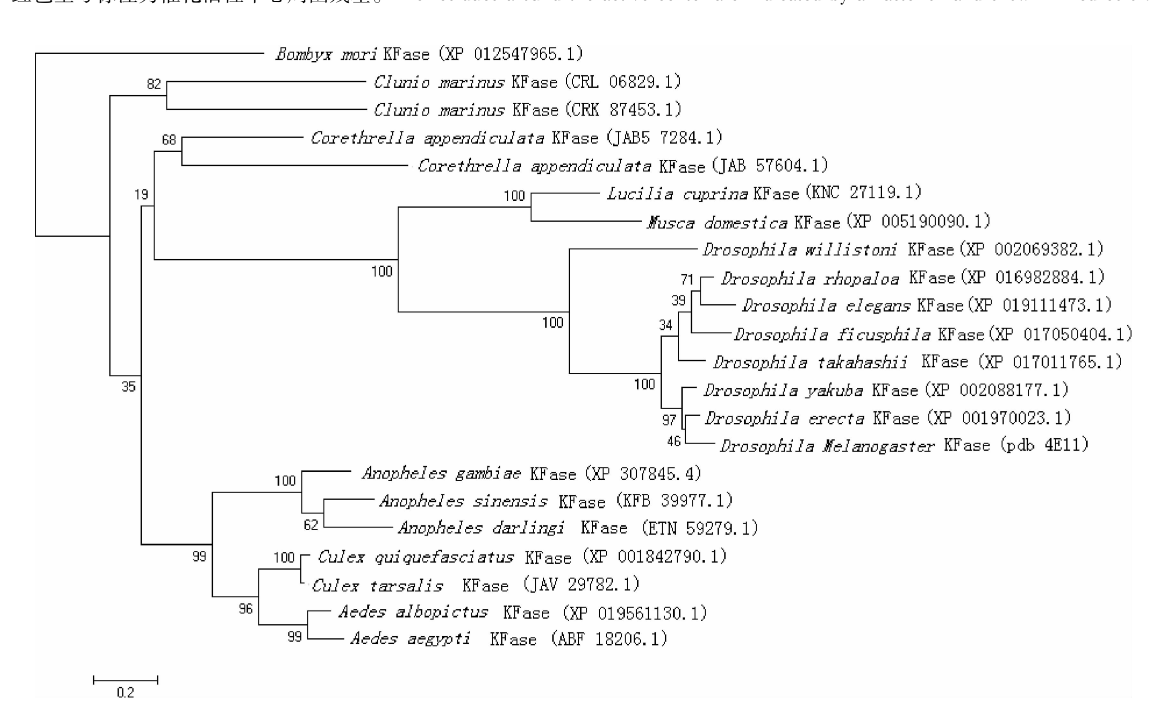


图 3 冈比亚按蚊 KFase 与 BLASTP 检索的同源蛋白的进化树分析（最大似然法）

Fig. 3 Phylogenetic analysis of KFase of *Anopheles gamiae* and homologous proteins searched by BLASTP (maximum likelihood method)

标尺示遗传距离；Bootstrap 的百分率在进化树的节点上标出。The scale bar indicates the genetic distance, and the percentage of bootstrap is indicated on branches.

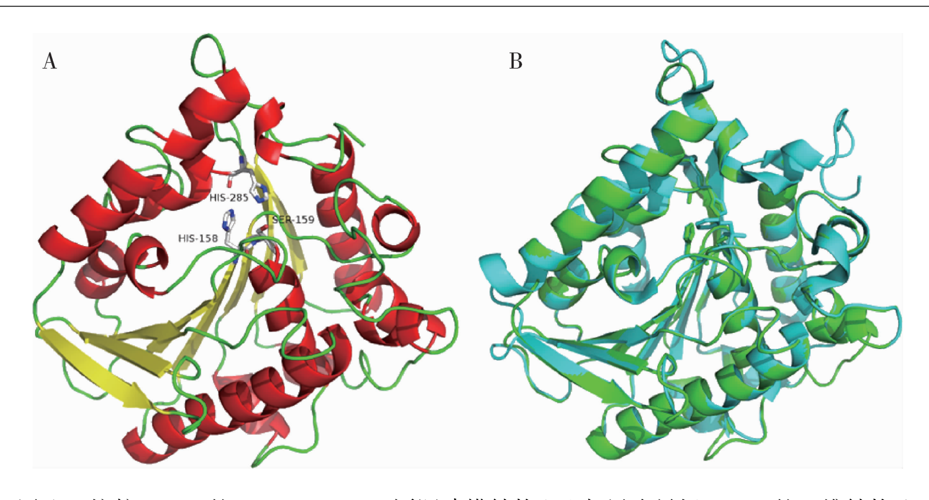


图 4　冈比亚按蚊 KFase 的 SWISS-MODEL 同源建模结构（A）与黑腹果蝇 KFase 的三维结构比对（B）

Fig. 4　Model of KFase of *Anopheles gambiae* got by SWISS-MODEL (A) and the 3D alignment with KFase of *Drosophila melanogaster* (B)

绿色为冈比亚按蚊 KFase 三维结构，蓝色为黑腹果蝇 KFase 三维结构。The 3D structure of *An. gambiae* is shown in green color and that of *D. melanogaster* in blue color.

2.3　冈比亚按蚊 KFase 的动力学模拟

SWISS-MODEL 构建的冈比亚按蚊 KFase 结构模型在水环境中 20 ns 的动力学模拟后，蛋白结构发生了一定程度的形变，但最核心的部分仍保持稳定，主要的 α-螺旋及 β-片层结构均未遭受破坏，相对位置变化较小；组成活性中心的 SER159，ASP253 和 HIS285 依然彼此靠近，因此活性中心的结构也未遭受破坏，如图 5 所示，图 5（A）为动力学模拟前的同源建模模型，图 5（B）为模拟在水环境中 50 ns 动力学过程后的 KFase 结构模型。图 6 中可观察到分子动力学模拟过程中分子势能稳定。

2.4　底物与酶复合物动力学模拟

冈比亚按蚊 KFase 和底物 NFK 的复合物模型在 20 ns 的动力学过程后，达到了较为稳定的结合态（图 7）。图 7（A）为能量最小化过程中的分子势能变化，图中显示在前 200 ps 体系能量迅速下降，200 ps 后体系能量趋于稳定。图 7（B）为动力学模拟过程分子势能变化，图 7（C）为动力学模拟过程中的分子动能变化，图中显示经过复合物模型体系在 20 ns 的动力学模拟过程中分子势能和分子动能稳定。

2.5　随机森林方法对小分子化合物数据库的筛选

经 MOLD2 处理，779 个描述符筛得 171 个描述符。随后利用 F-score 进行特征选择。最终从 171 个描述符中得到 49 个最佳特征描述符，用于对小分子化合物数据库进行随机森林筛选。对含有超过 70 万个化合物的数据库进行随机森林筛选，取大于 400 投票的数据，计算 5 次，取交集，共有 405 个化合物。随后用 AutoDock Vina 化合物对这 405 个化合物进行分子对接，每个化合物对接 3 次，取亲和能最低者，进行排序。

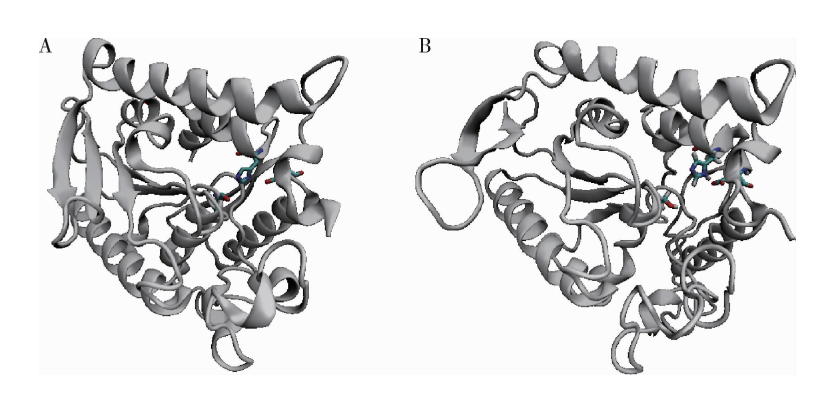


图 5　冈比亚按蚊 KFase 结构模型动力学模拟前（A）和后（B）分子模型的构象

Fig. 5　Molecular structure of KFase of *Anopheles gambiae* model before (A) and after (B) molecular dynamics simulation

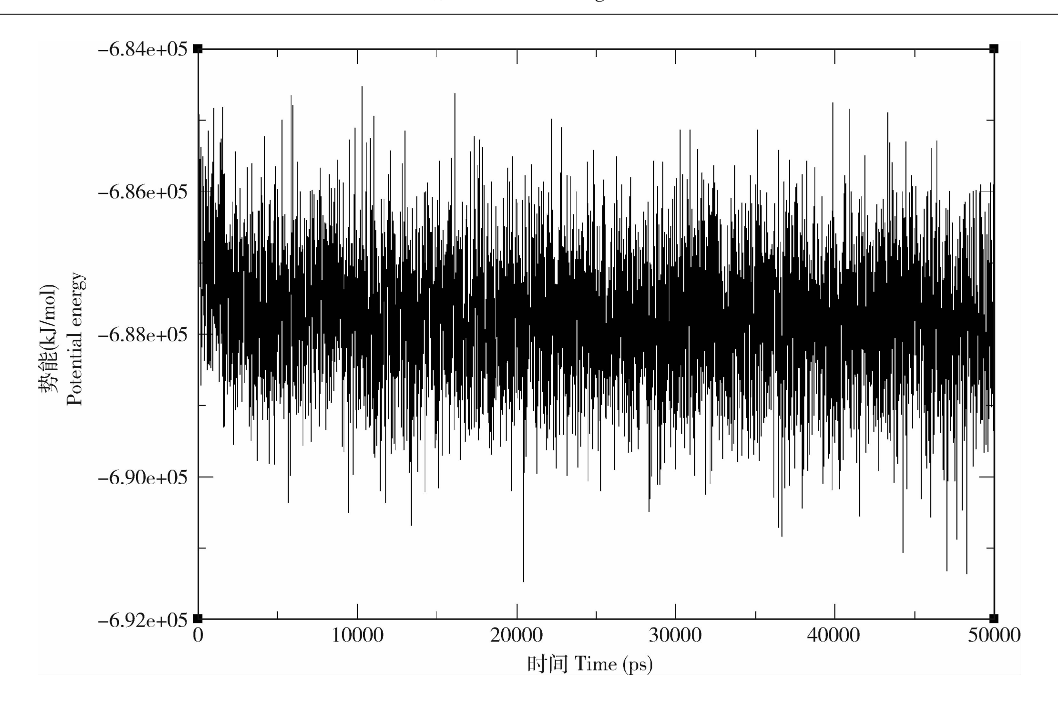


图 6 动力学模拟过程中冈比亚按蚊 KFase 分子势能的变化

Fig. 6 Changes in the potential energy of KFase of *Anopheles gambiae* during molecular dynamics simulation

编号 ZINC02711341（N-(2,4-diketo-1H-pyrimidin-6-yl)-2-fluoro-benzamide）, ZINC26420338（3-(4-fluorophenyl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylic acid）和 ZINC38548901（N-(2-oxo-2,3-dihydro-1H-imidazo[4,5-b]pyridin-5-yl)-succinamic acid）的物质在半柔性对接的虚拟筛选中，其亲和能最低，同时构象最佳。将该 3 种化合物作为配体，分别与酶形成复合物用于动力学模拟。

2.6 配体与酶复合物的动力学模拟

编号 ZINC26420338 的小分子化合物与冈比亚按蚊 KFase 复合物（图 8：A）利用刚性对接计算亲和能时，显现出了 -8.7 kcal/mol 的亲和能，推测其为冈比亚按蚊 KFase 的潜在竞争性抑制剂。在模拟 10 ns 的动力学过程中该化合物与冈比亚按蚊 KFase 的活性中心充分结合，同时分子势能稳定（图 9：A）。

编号 ZINC02711341 的小分子化合物与冈比亚按蚊 KFase 复合物（图 8：B）利用刚性对接计算亲和能时，显现出了 -9.0 kcal/mol 的亲和能，推测其为冈比亚按蚊 KFase 的潜在竞争性抑制剂。在模拟 10 ns 的动力学过程中该化合物与冈比亚按蚊 KFase 的活性中心充分结合，同时分子势能稳定（图 9：B）。

编号 ZINC38548901 的小分子化合物与冈比亚按蚊 KFase 复合物（图 8：C）利用刚性对接计算亲和能时，显现出了 -8.9 kcal/mol 的亲和能，推测其为冈比亚按蚊 KFase 的潜在竞争性抑制剂。在模拟 10 ns 的动力学过程中该化合物与冈比亚按蚊 KFase 的活性中心结合，同时分子势能稳定（图 9：C）。

3 讨论

随着计算机硬件制造技术的不断进步，计算机的应用软件技术也得到长足的发展。结合随机森林算法进行机器学习和虚拟筛选，并利用同源建模、动力学模拟和分子对接等手段，除了用于对目标蛋白质的抑制剂进行虚拟筛选外，还可用于激动剂等其他类型配体的虚拟筛选。虽然随机森林筛选，用于生物信息学领域进行虚拟筛选尚不十分普遍，但相关筛选技术已较为成熟。如同源建模技术由于大量蛋白质结构被解析，并收集在蛋白质结构数据库 PDB（http://www.rcsb.org），获得大量的结构模板；以经典分子力学为基础的动力学模拟技术，实现在分子水平阐释分子间的相互作用机制；分子对接技术在药物设计等领域成为常用的技术。Kain 等（2013）以驱蚊胺受体为靶标的虚拟筛选，用到了多种方法和技术，包括基于前进序列选择（sequential-forward-selection）算法的机器学习和虚拟筛选，找到多种小分子化合物可以作为新型蚊虫驱避剂。赵瑞等（2015）以 5-羟色胺转运体为受体靶标，运用生物筛选结合虚拟筛选进行了大量样本的高通量筛选，

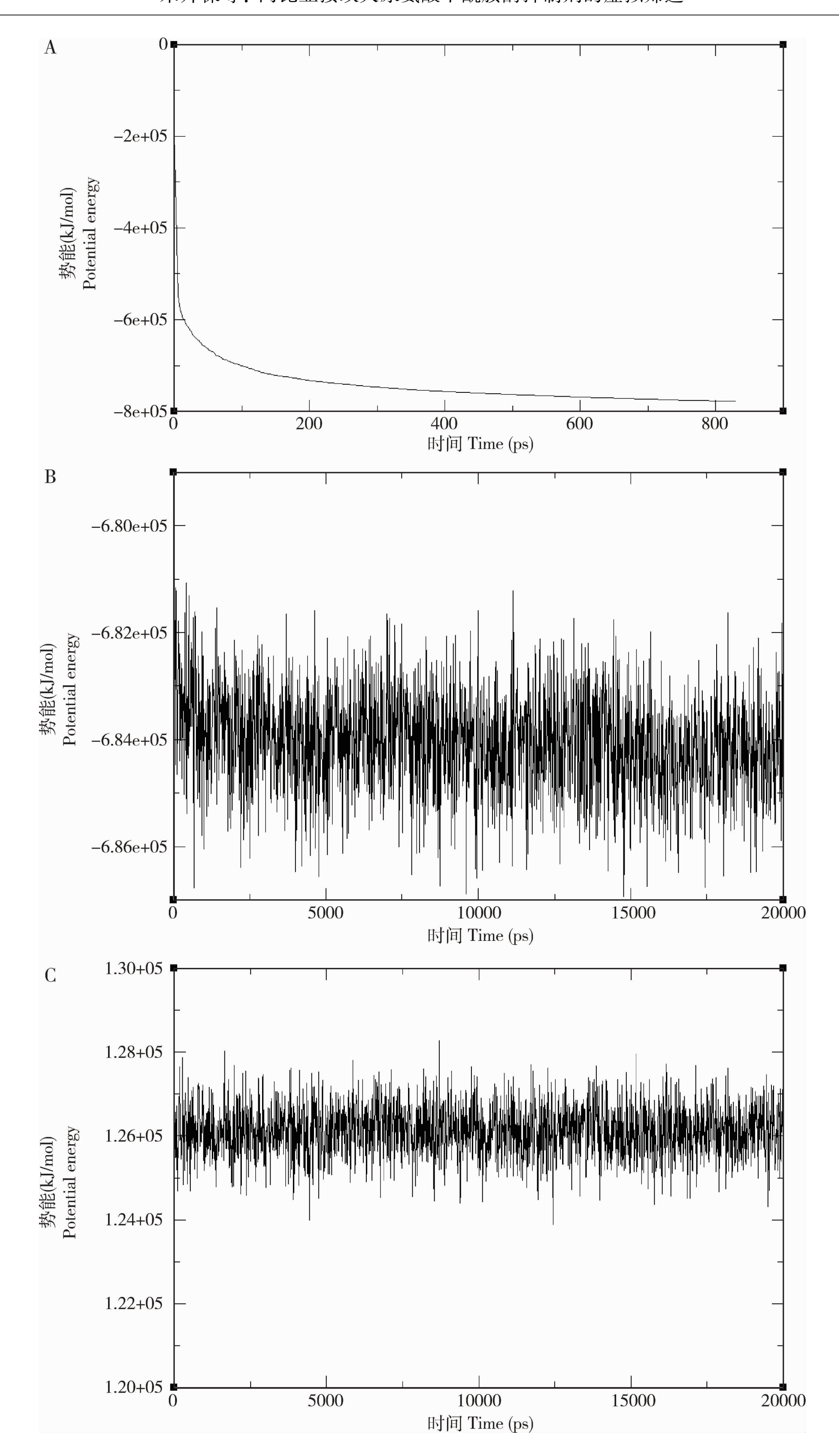


图 7　NFK 与冈比亚按蚊 KFase 的复合物在动力学模拟过程中的能量变化

Fig. 7　Energy changes of the complex of NFK and KFase of *Anopheles gambiae* during molecular dynamics simulation

A：能量最小化过程中分子的势能变化 Molecular potential energy changes in energy minimization；B：动力学模拟过程中分子的势能变化 Molecular potential energy changes in dynamics simulation；C：动力学模拟过程中分子的动能变化 Molecular kinetic energy changes in dynamics simulation.

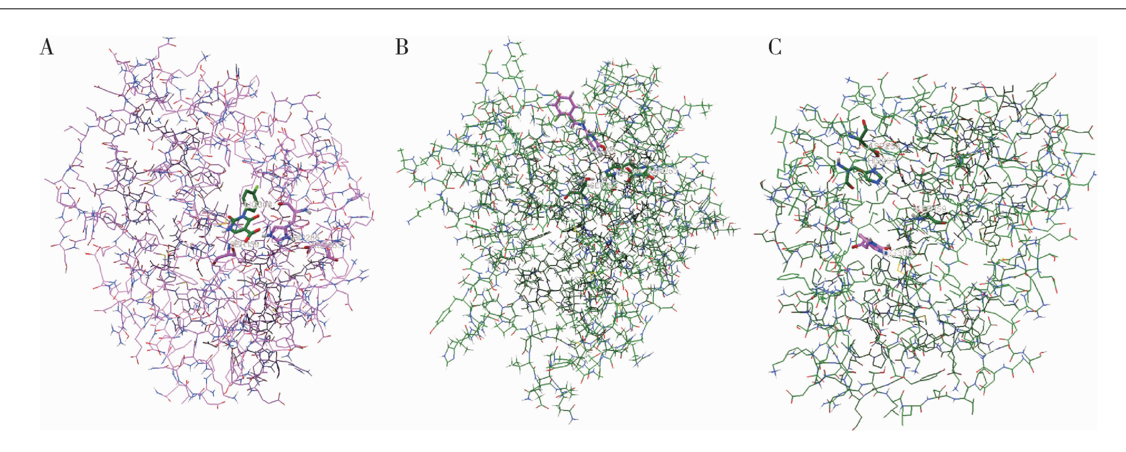


图 8　小分子化合物 ZINC26420338(A)，ZINC02711341(B)和 ZINC38548901(C)与冈比亚按蚊 KFase 复合物的结构模型
Fig. 8　Complex structures of small-molecular compounds ZINC26420338 (A), ZINC02711341 (B) and ZINC38548901 (C) with KFase of *Anopheles gambiae*

成功筛选出了 30 个达到预期抑制作用的化合物。汪滢等(2015)则以 HIV-1 蛋白酶为受体靶标,结合分子对接等多种方法建立了兼顾计算效率和预测准确度的筛选方法,成功筛选出了与美国 FDA 批准上市的药物 Atazanavir 具备相似性,但结构更新颖的的抑制剂。

由于蚊虫的 KFase 晶体结构尚未得到解析,但根据同源比对,冈比亚按蚊 KFase 序列与黑腹果蝇 KFase 的序列有较高相似性,尤其是催化反应中心附近的氨基酸残基具有高度的相似性,因此可以选择黑腹果蝇 KFase 的三维结构作为模板对冈比亚按蚊 KFase 进行建模。通过对同源建模后的冈比亚按蚊 KFase 进行拉氏构象图分析,大部分氨基酸残基的二面角位于构象允许区内,并且分析动力学模拟过程中的蛋白质形变程度和能量变化情况,说明此模型在水环境中是稳定的,说明选择黑腹果蝇 KFase 的三维结构作为模板是合适的。

ZINC 数据库中 3-(4-fluorophenyl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylic acid(编号 ZINC26420338)、N-(2,4-diketo-1H-pyrimidin-6-yl)-2-fluoro-benzamide(编号 ZINC02711341)和 N-(2-oxo-2,3-dihydro-1H-imidazo[4,5-b]pyridin-5-yl)-succinamic acid(编号 ZINC38548901)是本次虚拟筛选的最佳结果,在分子对接中体现了与冈比亚按蚊 KFase 较好的对接效果。在利用刚性对接计算亲和能时,亲和能分别为 －8.7, －9.0 和 －8.9 kcal/mol。这些化合物可能是冈比亚按蚊犬尿氨酸甲酰胺酶的潜在竞争性抑制剂。同时由于相对分子质量较小,可能具备一定的挥发性。然而它们对冈比亚按蚊 KFase 的抑制作用和是否有杀蚊效果还需做更深入的试验验证。另外,该 3 种化合物的应用目前还没有文献报道,其理化性质尤其是毒理和药理分析有待深入研究。

参考文献(References)

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E, 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1: 19－25.

Arnold K, Bordoli L, Kopp J, Schwede T, 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22(2): 195－201.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, 42 (Web Server issue): W252－W258.

Breiman L, 2001. Random forests. *Machine Learning*, 45(1): 5－32.

Chen YW, Lin CJ, 2006. Combining SVMs with various feature selection strategies. In: Guyon I, Nikravesh M, Gunn S, Zadeh LA eds. Feature Extraction. Studies in Fuzziness and Soft Computing, Vol. 207. Springer, Berlin, Heidelberg. 315－324.

Garcia GE, Wirtz RA, Barr JR, Woolfitt A, Rosenberg R, 1998. Xanthurenic acid induces gametogenesis in *Plasmodium*, the malaria parasite. *J. Biol. Chem.*, 273(20): 12003－12005.

Han Q, Robinson H, Li J, 2012. Biochemical identification and crystal structure of kynurenine formamidase from *Drosophila melanogaster*. *Biochem. J.*, 446(2): 253－260.

Hong HX, Xie Q, Ge WG, Qian F, Fang H, Shi LM, Su ZQ, Perkins R, Tong WD, 2008. Mold^2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *J. Chem. Inf. Model.*, 48(7): 1337－1344.

Jaghoori MM, Bleijlevens B, Olabarriaga SD, 2016. 1001 Ways to run AutoDock Vina for virtual screening. *J. Comput. Aided Mol. Des.*, 30(3): 237－249.

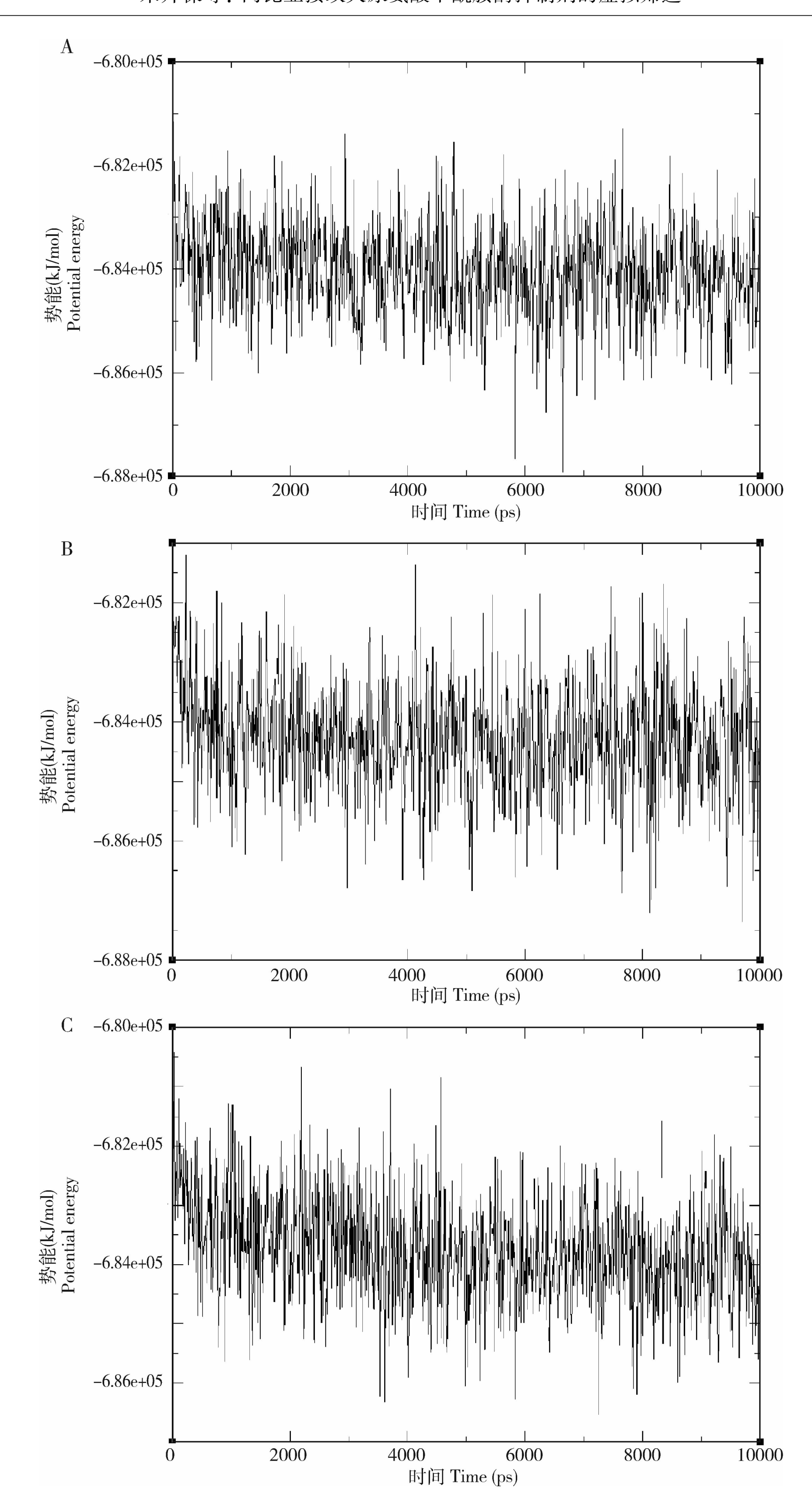


图 9　小分子化合物 ZINC26420338(A),ZINC02711341(B)和 ZINC38548901(C)与冈比亚按蚊 KFase 复合物的结构模型的总能量

Fig. 9　Total energy of the complex structures of small-molecular compounds ZINC26420338 (A), ZINC02711341 (B) and ZINC38548901 (C) with KFase of *Anopheles gambiae*

Kain P, Boyle SM, Tharadra SK, Guda T, Pham C, Dahanukar A, Ray A, 2013. Odour receptors and neurons for DEET and new insect repellents. *Nature*, 502(7472): 507 – 512.

Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T, 2009. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.*, 37(Database issue): D387 – D392.

Kutzner C, Páll S, Fechner M, Esztermann A, De Groot BL, Grubmüller H, 2015. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J. Comput. Chem.*, 36(26): 1990 – 2008.

Liu YF, Zhang J, Li H, 2002. Application of carbamate compounds. *Chem. Bull.*, 65(3): 167 – 173. [刘毅锋, 张娟, 李华, 2002. 氨基甲酸酯类化合物的应用. 化学通报, 65(3): 167 – 173]

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ, 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.*, 30(16): 2785 – 2791.

Nagai T, Serrano AE, Nagayama F, 1994. Purification and properties of skipjack liver tryptophan hydroxylase. *Fish. Sci.*, 60(4): 445 – 448.

Silva AWSD, Vranken WF, 2012. ACPYPE – AnteChamber PYthon Parser interfacE. *BMC Res. Notes*, 5(1): 367.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.*, 30(12): 2725 – 2729.

Tanchuk VY, Tanin VO, Vovk AI, Poda G, 2015. A new scoring function for molecular docking based on AutoDock and AutoDock Vina. *Curr. Drug Discov. Technol.*, 12(3): 170.

Trott O, Olson AJ, 2010. Software news and update AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, 31(2): 455 – 461.

Wang J, Wang W, Kollman PA, Case DA, 2006. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model*, 25(2): 247 – 260.

Wang Y, Tang GR, Liu SN, 2015. Discovery of new HIV-1 protease inhibitors by integrating molecular similarity, pharmacophore and docking methods. *Chin. J. Bioinform.*, 13(4): 244 – 250. [汪滢, 唐国荣, 刘澍楠, 2015. 结合分子相似性、药效团和分子对接筛选新的 HIV-1 蛋白酶抑制剂. 生物信息学, 13(4): 244 – 250]

WHO, 2015. World Malaria Report 2015. Available at: http://apps.who.int/iris/bitstream/10665/200018/1/9789241565158_eng.pdf.

Xie JY, Wang CX, Jiang S, Zhang Y, 2010. Feature selection method combing improved F-score and support vector machine. *J. Comput. Appl.*, 30(4): 993 – 996. [谢娟英, 王春霞, 蒋帅, 张琰, 2010. 基于改进的 F-score 与支持向量机的特征选择方法. 计算机应用, 30(4): 993 – 996]

Zhao R, Fang JS, Liu AL, Du GH, 2015. Development of HTS model on SERT inhibitors combined biological screening model with HTVS. *Acta Pharm. Sin.*, 50(9): 1116 – 1121. [赵瑞, 方坚松, 刘艾林, 杜冠华, 2015. 虚拟筛选技术辅助 SERT 抑制剂高通量筛选模型的建立及应用. 药学学报, 50(9): 1116 – 1121]

（责任编辑：赵利辉）